

REC'D 07 JAN 2005

WIPO



INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

PRIORITY DOCUMENT

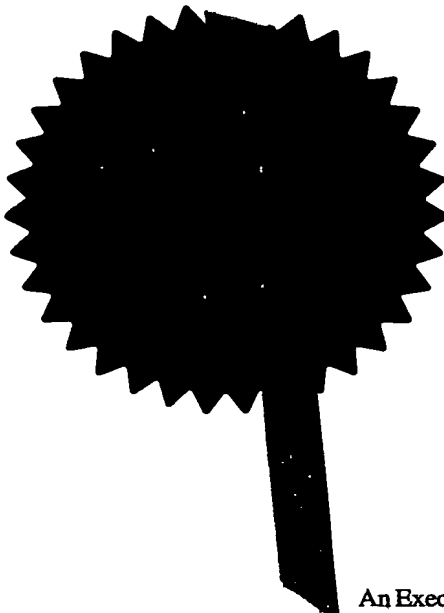
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



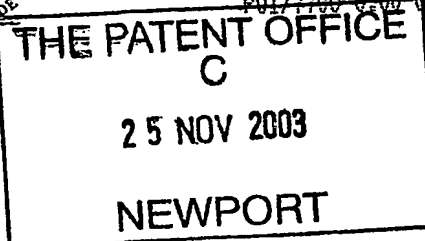
Signed

Dated

22 December 2004



25NOV03 E854705.4 D02884
P01/7700 0.00 0327338.0



The Patent Office

Cardiff Road
Newport
South Wales
NP10 8QQ

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

1. Your reference	P339851/CMU/RTH/RMC		
2. Patent application number (The Patent Office will fill in this part)	0327338.0		
3. Full name, address and postcode of the or of each applicant (underline all surnames)	The Queen's University of Belfast University Road Belfast, BT7 1NN		
Patents ADP number (if you know it) If the applicant is a corporate body, give the country/state of its incorporation	889659006 United Kingdom		
4. Title of the invention	"Assay"		
5. Name of your agent (if you have one) "Address for service" in the United Kingdom to which all correspondence should be sent (including the postcode)	Murgitroyd & Company Scotland House 165-169 Scotland Street Glasgow G5 8PL		
Patents ADP number (if you know it)	1198015 /		
6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and (if you know it) the or each application number	Country	Priority application number (if you know it)	Date of filing (day / month / year)
7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application	Number of earlier application	Date of filing (day / month / year)	
8. Is a statement of inventorship and of right to grant of a patent required in support of this request? (Answer 'Yes' if: a) any applicant named in part 3 is not an inventor, or b) there is an inventor who is not named as an applicant, or c) any named applicant is a corporate body. See note (d))	Yes		

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form

Description

47

Claim(s)

10

Abstract

Drawing(s)

4

+ 4 RM.

10. If you are also filing any of the following, state how many against each item.

Priority documents

Translations of priority documents

Statement of inventorship and right to grant of a patent (*Patents Form 7/77*)

Request for preliminary examination and search (*Patents Form 9/77*)

Request for substantive examination (*Patents Form 10/77*)

Any other documents
(*please specify*)

11. I/We request the grant of a patent on the basis of this application.

Signature

Murgitroyd & Co

Date

24 November 2003

Murgitroyd & Company

12. Name and daytime telephone number of person to contact in the United Kingdom

MALCOLM MAIN

0141 307 8400

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

- If you need help to fill in this form or you have any questions, please contact the Patent Office on 08459 500505.
- Write your answers in capital letters using black ink or you may type them.
- If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.
- If you have answered 'Yes' Patents Form 7/77 will need to be filed.
- Once you have filled in the form you must remember to sign and date it.
- For details of the fee and ways to pay please contact the Patent Office.

1 **"Detection of Protein Interactions"**

2

3 Field of the Invention

4

5 The present invention relates to a method of
6 detecting interactions between macromolecules. In
7 particular, but not exclusively, the invention
8 relates to a method of detecting protein
9 interactions using fluorescence.

10

11 Background to the Invention

12

13 Protein to protein interactions play a key role in
14 many biological processes including the assembly of
15 enzymes, protein homo/hetero-oligomers, regulation
16 of intracellular transport, gene expression,
17 receptor-ligand interactions, entry of pathogens
18 into the cell and the action of small molecules or
19 drugs.

1
2 Identification and characterisation of
3 macromolecular interactions can be performed using
4 co-immunoprecipitation from cell lysates and
5 solubilised membranes. However, this technique
6 requires specific antibodies for both capture and
7 identification of proteins and may further require
8 the use of detergent to disrupt interactions.

9
10 More recently non-invasive techniques have been
11 developed to determine protein to protein
12 interactions.

13
14 Such non-invasive techniques were pioneered by the
15 yeast two hybrid method which is based on
16 complementation of a split yeast nuclear
17 transcription factor. The yeast two hybrid method
18 utilises a mammalian bait protein connected to a
19 yeast DNA-binding domain. This bait protein is used
20 to determine which prey proteins are able to bind to
21 the bait protein from a mixture of prey proteins.
22 The mammalian prey protein is connected to a yeast
23 transcription activation domain. When the mammalian
24 bait and prey proteins interact, the yeast DNA-
25 binding and transcription activating domains are
26 brought together. The DNA binding domain can bind
27 to the yeast DNA and the transcription activating
28 domain is then suitably located to trigger the
29 expression of a reporter gene encoding an enzyme
30 which in turn can catalyse the production of a
31 coloured product within the yeast cells thus

1 indicating a successful interaction of bait with
2 prey.

3

4 The use of yeast expression systems to identify
5 mammalian protein-to-protein interaction suffers
6 from a number of flaws. Certain post-translational
7 modifications, that are normally critical to
8 mammalian protein interactions, cannot be achieved
9 by yeast cells. For example, tyrosine
10 phosphorylation is key to many mammalian
11 intracellular protein binding events involved in
12 signal transduction. However, the yeast genome
13 contains no tyrosine kinase genes so
14 phosphotyrosine-dependent protein interactions
15 cannot be accessed in yeast two hybrid studies.

16

17 Furthermore, in yeast two hybrid screening the
18 protein complex must be able to translocate to the
19 nucleus to cause expression of the reporter gene or
20 cause downstream events to trigger the expression of
21 a reporter gene. Thus proteins that are excluded
22 from the yeast nucleus will not be accessible to
23 this screening method.

24

25 Further methods such as protein complementation and
26 the split ubiquitin method utilise similar
27 underlying concepts to the yeast two hybrid method
28 in that the interaction of two proteins (a bait and
29 prey protein) act to express a reporter protein, the
30 reporter gene allowing the interaction event to be
31 visualised as a detectable signal.

32

1 Such methods which utilise the expression of a
2 reporter enzyme to produce a detectable signal
3 suffer from the disadvantage that the location of
4 the protein complexes being detected cannot be
5 accurately visualised in the cell.

6
7 Recently the technique of fluorescence energy
8 transfer (FRET) has been used to determine protein
9 to protein interactions. In this technique the
10 interaction of two fluorophores indicates their
11 close spatial proximity. For protein to protein
12 interaction monitoring the addition of an absorbing
13 moiety to one protein partner is complemented by the
14 addition of a second fluorescing moiety to the
15 second binding partner. Provided the emission
16 spectrum of the absorbing moiety overlaps the
17 excitation spectrum of the fluorescing moiety and
18 both moieties are within 100Å of each other FRET
19 will occur. Mutations in the sequence of green
20 fluorescent protein (GFP) from the jellyfish
21 *Aequorea victoria* have been studied and shown to
22 cause variations in the spectral emission of GFP
23 giving rise to variants of GFP such as Yellow
24 Fluorescent Protein (YFP), as well as cyan (CFP) and
25 blue (BFP) fluorescing variants. This technique uses
26 fluorescent energy transfer between these colour
27 variants of GFP which are fused to interacting
28 proteins to determine protein to protein
29 interaction. Using this method, when the two GFP
30 derived fluorophores are brought into close
31 proximity, energy transfer between the fluorescent
32 variants occurs and changes in fluorescence

1 emissions are detected. Unfortunately, this method
2 requires overexpression of the GFP fusion proteins
3 to allow quantification of the small changes in
4 fluorescence. Related methods to FRET require the
5 use of irreversible photobleaching (FRAP) or
6 expensive instruments capable of measuring
7 fluorescence lifetime imaging (FLIM).

8
9 As a preliminary to the current experiments it was
10 shown that green fluorescent protein can be
11 engineered to add amino acid residues at particular
12 regions in the GFP sequence whilst fluorescence is
13 retained. Further, it has been shown in Hu, CD,
14 Chinenov, Y. and Kerppola, T. K. (2002). *Mol. Cell.*
15 *9*, 789-798 that using recombinant DNA technology
16 specific Yellow Fluorescent Protein (YFP) fragments
17 covalently fused to peptide sequences, which are
18 capable of interacting with each other can
19 reconstitute a fluorophore when the YFP fragments
20 are brought together, such that the peptide
21 sequences could interact.

22
23 Further, it has recently been shown that
24 fluorescence can be generated following the
25 functional association of two separate fragments of
26 the GFP molecule (hapto-GFPs) when driven by the
27 interaction of a pair of proteins fused via a linker
28 to the new C' and N' termini of the hapto-GFPs.
29 (Ghosh et al, (2000); Hu et al, (2002)).

30
31 However, the above methods suffer from the
32 disadvantage that functional association of

1 fluorescent fragments is limited by the constraints
2 of stereochemistry imposed on the fragments by the
3 bait and prey proteins' association. If the fusion
4 termini of the interacting partners are widely
5 separated, productive association of the haptGFPs
6 will not occur and no signal will be generated to
7 indicate the interaction between the bait and prey
8 peptides.

9

10 The present inventors have overcome a number of
11 problems of the prior art.

12

13 Summary of the Invention

14

15 According to a first aspect of the present invention
16 there is provided a protein interaction system said
17 system comprising a first construct which encodes a
18 first fragment of fluorescent protein, a first bait
19 peptide and a linker portion encoding at least 5
20 amino acid residues interposed between the first
21 fragment and the bait peptide and a plurality of
22 second constructs encoding a second fragment of
23 fluorescent protein, a prey peptide and a linker
24 portion encoding at least 5 amino acid residues
25 interposed between the second fragment and the prey
26 peptide and on interaction of the bait and a prey
27 peptide the first and second fragments of the
28 fluorescent protein complement each other such that
29 functional association of the first and second
30 fragments promotes fluorescence, wherein at least
31 two of the prey proteins have different amino acid
32 sequences.

1

2 Preferably all the prey proteins have different
3 amino acid sequences.

4 Preferably the first and / or second construct
5 comprises a linker portion which encodes between 15
6 to 100 amino acid residues.

7

8 Preferably the linker of the first and / or second
9 construct is comprised of substantially hydrophillic
10 amino-acid residues.

11

12 More preferably the linker of the first and / or
13 second construct is comprised of multiples of a
14 pentapeptide sequence such as glycyl-glycyl-glycyl-
15 glycyl-serine.

16

17 More preferably the linker of the first and / or
18 second construct is greater than 20 amino acids,
19 more preferably greater than 25 amino acids, more
20 preferably greater than 30 amino acids, more
21 preferably greater than 35 amino acids, even more
22 preferably greater than 40 amino acids, even more
23 preferably greater than 50 amino acids and yet more
24 preferably greater than 55 amino acids in length.

25

26 Preferably, the linker of the first and / or second
27 construct encodes up to 60 amino acids.

28

29 Where the peptides joined to the linkers are rod
30 like structures and the peptides interact with each
31 other with favourable topology of interaction, ie
32 the peptides interact such that the fragments of

1 fluorescent protein are brought into close proximity
2 with each other, short linker lengths are sufficient
3 to allow screening for interaction partners. For
4 example, short linkers could be used to screen a
5 library of DNA binding proteins which from previous
6 studies are known to be rod like in structure.

7
8 However, linker lengths between 15 to 100 amino
9 acids are advantageous over shorter linker lengths
10 as they allow bulkier peptides being tested for
11 interaction to be conjoined to the first and second
12 fragments of the fluorescent protein without the
13 peptides being tested placing constraints on the
14 functional association of the fluorescent proteins
15 due to stereochemical hindrance. Such longer
16 linkers are also advantageous to study small peptide
17 pairs that have an unfavourable topology of
18 interaction such as is found in an anti-parallel
19 complex (hapto-GFP- $N^1 \rightarrow C^1$:binding to $C^2 \rightarrow N^2$ -hapto-
20 GFP) i.e. functional association of the interacting
21 peptides causes the fluorescent fragments to be
22 orientated such that they are directed away from
23 each other in space.

24
25 Any fluorescent protein may be used in the
26 invention. However, in a preferred embodiment the
27 fragments of fluorescent protein are fragments of
28 green fluorescent protein, mutants or variants
29 thereof.

30
31 More preferably the fluorescent protein is the
32 humanised form of a fluorescent protein, e.g.

1 Enhanced Green Fluorescent Protein (EGFP) or a
2 variant thereof.

3
4 Variants include peptides in which individual amino
5 acids are substituted by other amino acids which are
6 closely related as understood in the art, for
7 example, substitution of one hydrophobic residue
8 such as isoleucine, valine, leucine or methionine
9 for another, or the substitution of one polar
10 residue for another, such as arginine for lysine,
11 glutamic for aspartic acid or glutamine for
12 asparagine.

13
14 In a humanised nucleotide sequence one or more of
15 the codons in the sequence are altered such that for
16 the amino acid being encoded, the codon used is that
17 which most frequently appears in humans. This is
18 advantageous as the humanised fluorescent protein
19 construct e.g. (EGFP) has maximised expression
20 levels and rate of fluorophore formation in mammalian
21 cells. This makes detection of fluorescence,
22 produced by fragments of fluorescent proteins
23 (fluorogenic fragments) which functionally associate
24 with each other, easier to determine.

25
26 In a second aspect, there is provided a library of
27 constructs encoding a fragment of fluorescent
28 protein, a peptide and a linker portion of at least
29 5 amino acids interposed between said fragment and
30 peptide wherein said fragment of fluorescent protein
31 is capable of functional association with a
32 complementary fragment of fluorescent protein such

1 that on functional association of said fragments
2 fluorescence is enabled wherein the library of
3 constructs encodes a plurality of different
4 peptides.

5
6 Each member of the library encodes or provides a
7 different peptide fused to a fragment of fluorescent
8 protein via a linker. The peptides can be small
9 peptides of differing amino acid sequence, for
10 example nonomers, comprising different amino acid
11 compositions or the same overall composition but
12 with the amino acids present in a different order.
13 Alternatively the peptides may be full size proteins
14 obtained from a cDNA library.

15
16 Preferably the constructs of the library comprise a
17 linker portion which encodes between 15 to 100 amino
18 acid residues.

19
20 Preferably the linker is comprised of substantially
21 hydrophillic amino-acid residues.

22
23 More preferably the linker is comprised of multiples
24 of a pentapeptide sequence such as glycyl-glycyl-
25 glycyl-serine.

26
27 More preferably the linker portion is greater than
28 20 amino acids, more preferably greater than 25
29 amino acids, more preferably greater than 30 amino
30 acids, more preferably greater than 35 amino acids,
31 even more preferably greater than 40 amino acids,
32 even more preferably greater than 50 amino acids and

1 yet more preferably greater than 55 amino acids in
2 length.

3

4 Preferably, the linker encodes up to 60 amino acids.

5

6 The invention further provides in a third aspect a
7 library of polypeptides, each polypeptide comprising
8 a fragment of fluorescent protein, a peptide and a
9 linker portion of at least 5 amino acid residues
10 interposed between the fragment and the peptide of
11 the polypeptide.

12

13 Unless the context demands otherwise, the term
14 peptide, polypeptide and protein are used
15 interchangeably to refer to amino acids in which the
16 amino acid residues are linked by covalent peptide
17 bonds or alternatively (where post-translational
18 processing has removed an internal segment) by
19 covalent di-sulphide bonds, etc. The amino acid
20 chains can be of any length and comprise at least
21 two amino acids, they can include domains of
22 proteins or full-length proteins. Unless otherwise
23 stated the terms, peptide, polypeptide and protein
24 also encompass various modified forms thereof,
25 including but not limited to glycosylated forms,
26 phosphorylated forms etc.

27

28 Polypeptides may be made synthetically or
29 recombinantly using techniques which are widely
30 available in the art.

31

32

1 In preferred embodiments, the fragments of
 2 fluorescent protein (fluorogenic fragments) are
 3 generatable through the introduction of a split
 4 point between the amino acids at positions 157 and
 5 158, or (in a second embodiment) between the amino
 6 acids at positions 172 and 173 of the humanised form
 7 of Green Fluorescent Protein (SEQ ID NO 1).

8
 9 SEQ ID NO 1 - EGFP (Clontech Inc.) [Genebank
 10 Accession number gb:AAB02574 gi 1377912]:
 11 1 mvskgeelft gvpilveld gdvngkhkfsv sgegegdaty
 12 41 gkltlkfict tgklpvwpwt lvtlttygvq cfsrypdhmk
 13 81 qhdffksamp egyvqertif fkddgnyktr aevkfegdtl
 14 121 vnrielkgid fkedgnilgh kleynynshn vyimadkqkn
 15 161 gikvnfkirh niedgsvqla dhyqqntpig dgpvllpdnh
 16 201 ylstqsalsk dpnekrdhmv llefvtaagi tlgmdelyk

17
 18 The fluorogenic fragments generated by the
 19 introduction of a split point between the amino acid
 20 residues at positions 157 and 158, or between amino
 21 acid residues at positions 172 and 173, result in
 22 the production of hapto-EGFP^{1/157} and hapto-EGFP^{158/239},
 23 or hapto-EGFP^{1/172} and hapto-EGFP^{173/239}, respectively.

24
 25 Alternative split points are between residues 23/24,
 26 38/39, 50/51, 76/77, 89/90, 102/103, 116/117,
 27 132/133, 142/143, 190/191, 211/212, 214/215 of EGFP.

28
 29 Thus in preferred embodiments, the fragment
 30 comprises a fluorogenic fragment of amino acid
 31 residues 1 to 23, 1 to 38, 1 to 50, 1 to 76, 1 to
 32 89, 1 to 102, 1 to 116, 1 to 132, 1 to 142, 1 to

1 157, 1 to 172, 1 to 190, 1 to 211, 1 to 214, 24 to
2 239, 39 to 239, 51 to 239, 77 to 239, 90 to 239, 103
3 to 239, 117 to 239, 133 to 239, 143 to 239, 158 to
4 239, 173 to 239, 191 to 239, 212 to 239, or 215 to
5 239 of EGFP.

6
7 In one preferred embodiment a library of
8 polypeptides according to a further aspect of the
9 invention is provided wherein each member of the
10 library has a different peptide sequence fused to
11 the fragment of fluorescent protein via the linker
12 region.

13
14 A bait peptide is a sequence of two or more amino
15 acids, at least one domain of a protein or a full
16 length protein.

17
18 A prey peptide is a sequence of two or more amino
19 acids, at least one domain of a protein or a full
20 length protein.

21
22 The term interaction or interacting as used herein
23 means that two entities, for example, distinct
24 peptides, domains of proteins or complete proteins,
25 exhibit sufficient physical affinity to each other
26 so as to bring the two interacting entities
27 physically close to each other. An extreme case of
28 interaction is the formation of a chemical bond that
29 results in continual, stable proximity of the two
30 entities. Interactions that are based solely on
31 physical affinities, although usually more dynamic
32 than chemically bonding interactions, can be equally

1 effective at co-localising independent entities.
2 Physical affinities include, but are not limited to,
3 for example electrical charge differences,
4 hydrophobicity, hydrogen bonds, van der Waals force,
5 ionic force, covalent linkages, and combinations
6 thereof. The interacting entities may interact
7 transiently or permanently. Interaction may be
8 reversible or irreversible. In any event it is in
9 contrast to and distinguishable from natural random
10 movement of two entities. Examples of interactions
11 include specific interactions between antigen and
12 antibody, ligand and receptor etc.

13
14 In a fourth aspect of the invention there is
15 provided a protein interaction monitoring system,
16 said system comprising a first polypeptide
17 comprising a first fragment of fluorescent protein,
18 a bait peptide and a linker portion of at least 5
19 amino acid residues interposed between the first
20 fragment and the bait peptide and a plurality of
21 second polypeptides comprising a second fragment of
22 fluorescent protein, a prey peptide and a linker
23 portion of at least 5 amino acid residues interposed
24 between the second fragment and the prey peptide and
25 on interaction of the bait and a prey peptide the
26 first and second fragments of the fluorescent
27 protein complement each other such that functional
28 association of the first and second fragments
29 promotes fluorescence, wherein at least two of the
30 prey proteins have different amino acid sequences.
31

1 Preferably the linker portion of the first and / or
2 second polypeptide comprises between 15 to 100 amino
3 acid residues.

4

5 Preferably the linker of the first and / or second
6 polypeptide is comprised of substantially
7 hydrophillic amino-acid residues.

8

9 More preferably the linker of the first and / or
10 second polypeptide is comprised of multiples of a
11 pentapeptide sequence such as glycyl-glycyl-glycyl-
12 glycyl-serine.

13

14 More preferably the linker portion of the first and
15 / or second polypeptide is greater than 20 amino
16 acids, more preferably greater than 25 amino acids,
17 more preferably greater than 30 amino acids, more
18 preferably greater than 35 amino acids, even more
19 preferably greater than 40 amino acids, even more
20 preferably greater than 50 amino acids and yet more
21 preferably greater than 55 amino acids in length.

22

23 Preferably, the linker of the first and / or second
24 polypeptide comprises up to 60 amino acids.

25

26 According to a fifth aspect of the present
27 invention there is provided an assay method to
28 determine peptide to peptide interactions comprising
29 the steps of:

30

31 providing a first construct, said construct
32 encoding a first fragment of fluorescent

1 protein, a first bait peptide and a linker
2 portion of at least 5 amino acid residues
3 interposed between the first fragment and the
4 bait peptide;

5
6 providing a plurality of second constructs said
7 constructs encoding a second complementary
8 fragment of fluorescent protein, a prey peptide
9 and a linker portion of at least 5 amino acids
10 interposed between the second fragment and the
11 prey peptide wherein at least two constructs
12 encode different prey proteins;

13
14 expressing both constructs in the same cell;
15 and

16
17 detecting fluorescence produced in the cell.

18
19 Preferably all the second constructs encode
20 different prey proteins.

21
22 Preferably the first and / or second construct
23 comprises a linker portion which encodes between 15
24 to 100 amino acid residues.

25
26 Preferably the first and / or second linker is
27 comprised of substantially hydrophillic amino-acid
28 residues.

29
30 More preferably the first and / or second linker
31 encodes multiples of a pentapeptide sequence such as
32 glycyl-glycyl-glycyl-glycyl-serine.

1

2 More preferably the linker of the first and / or
3 second construct is greater than 20 amino acids,
4 more preferably greater than 25 amino acids, more
5 preferably greater than 30 amino acids, more
6 preferably greater than 35 amino acids, even more
7 preferably greater than 40 amino acids, even more
8 preferably greater than 50 amino acids and yet more
9 preferably greater than 55 amino acids in length.

10

11 Preferably, the linker of the first and / or second
12 construct encodes up to 60 amino acids.

13

14 In an embodiment of the assay the fluorescence
15 detected may be quantitatively determined such that
16 fluorescence produced by different cells or under
17 different conditions can be compared.

18

19 In one embodiment of the assay, the second construct
20 is provided as a member of a library of second
21 constructs wherein each member of the library
22 encodes a different prey peptide wherein at least
23 one second construct member of the library is
24 expressed in the same cell as the first construct
25 encoding the bait protein.

26

27 The assay can therefore be used to screen an
28 expression library to determine those peptides which
29 bind to a bait peptide.

30

31 There is also provided an assay to determine peptide
32 to peptide interactions comprising the steps of:

1
2 providing a first polypeptide comprising a
3 first fragment of fluorescent protein, a first
4 bait peptide and a linker portion of at least 5
5 amino acid residues interposed between the
6 first fragment and the bait peptide;

7
8 providing a plurality of second polypeptides
9 comprising a second fragment of fluorescent
10 protein which is complementary to the first
11 fragment of fluorescent protein, a prey peptide
12 and a linker portion of at least 5 amino acids
13 interposed between the second fragment and the
14 prey peptide wherein at least two second
15 polypeptides encode different prey proteins;

16
17 mixing the first polypeptide and second
18 polypeptide together; and

19
20 detecting whether fluorescence is produced.

21
22 Preferably the first and / or second polypeptide
23 linker portion comprises between 15 to 100 amino
24 acid residues.

25
26 Preferably the first and / or second polypeptide
27 linker is comprised of substantially hydrophillic
28 amino-acid residues.

29
30 More preferably the first and / or second
31 polypeptide linker is comprised of multiples of a

1 pentapeptide sequence such as glycyl-glycyl-glycyl-
2 glycyl-serine.

3

4 More preferably the first and / or second
5 polypeptide linker portion is greater than 20 amino
6 acids, more preferably greater than 25 amino acids,
7 more preferably greater than 30 amino acids, more
8 preferably greater than 35 amino acids, even more
9 preferably greater than 40 amino acids, even more
10 preferably greater than 50 amino acids and yet more
11 preferably greater than 55 amino acids in length.

12

13 Preferably, the first and / or second polypeptide
14 linker comprises up to 60 amino acids.

15

16 As detailed above the detected fluorescence can be
17 quantitatively measured.

18

19 In a particular example the assay method is
20 performed *in vitro*.

21

22 The assay method may further comprise the step of
23 determining the location of the fluorescence in the
24 cell. This is advantageous as it provides details
25 of not only if a protein to protein interaction is
26 occurring, but the location in the cell the
27 interaction is taking place, for example at the
28 membrane, in the cytoplasm, or in the nucleus.

29

30 In addition, the assay method may further comprise
31 the step of isolating the bait and / or prey peptide
32 encoded from the cell in which fluorescence has

1 resulted, for example isolating a cell using a
2 fluorescence activated cell sorting machine then
3 isolating and sequencing the interacting peptides.
4 The sequenced peptides can then be compared with
5 sequences (full length or partial) in a data bank so
6 as to identify or characterise the interacting
7 peptide isolated from the cell.

8
9 The sequences of the interacting peptides may
10 alternatively be inferred by cloning selected
11 fluorescent cells and subjecting the cloned cells to
12 PCR amplification and DNA sequencing. These
13 sequences can then be cloned into expression vectors
14 and the protein expressed and purified. The
15 purified protein can be further studied or used for
16 example in research.

17
18 In one embodiment, the assay method may further
19 comprise the process of determining the subcellular
20 dynamics of the peptide interactions visualised by
21 fluorescence observations of living cells to enable
22 spatio-temporal studies of peptide interactions
23 throughout all parts of the cell cycle.

24
25 In a sixth aspect, which enables spatio-temporal
26 studies, the invention provides an assay which
27 comprises the steps of providing a first construct
28 encoding a polypeptide comprising a first fragment
29 of fluorescent protein, a first bait peptide and a
30 linker portion of at least 5 amino acid residues
31 interposed between the first fluorogenic fragment
32 and the first bait peptide:

1
2 providing a second construct encoding a
3 polypeptide comprising a second fragment of
4 "fluorescent protein which is complementary to
5 said first fluorescent fragment, a second prey
6 peptide" and a linker portion interposed between
7 the second fluorogenic fragment and the second
8 prey peptide;
9
10 causing the expression of both constructs
11 within the same living cell; and
12
13 and observing the level of fluorescence
14 produced and its subcellular location in the
15 cell at a range of time points following co-
16 expression of both constructs.
17
18 Preferably the first and / or second construct
19 comprises a linker portion which encodes between 15
20 to 100 amino acid residues.
21
22 Preferably the linker of the first and / or second
23 construct is comprised of substantially hydrophillic
24 amino-acid residues.
25
26 More preferably the linker of the first and / or
27 second construct is comprised of multiples of a
28 pentapeptide sequence such as glycyl-glycyl-glycyl-
29 glycyl-serine.
30
31 More preferably the linker of the first and / or
32 second construct is greater than 20 amino acids,

1 more preferably greater than 25 amino acids, more
2 preferably greater than 30 amino acids, more
3 preferably greater than 35 amino acids, even more
4 preferably greater than 40 amino acids, even more
5 preferably greater than 50 amino acids and yet more
6 preferably greater than 55 amino acids in length.

7

8 Preferably, the linker of the first and / or second
9 construct encodes up to 60 amino acids.

10

11 In a seventh aspect, there is provided an assay for
12 estimating the maximum possible separation of the
13 fusion termini of the interacting peptides:

14

15 providing a first construct encoding a first
16 fragment of fluorescent protein, a first bait
17 peptide and a linker portion of at least 5
18 amino acid residues interposed between the
19 first fragment and the bait peptide;

20

21 providing a second construct encoding a second
22 fragment of fluorescent protein which is
23 complementary to said first fluorescent
24 fragment, a prey peptide and a library of
25 linkers of lengths ranging from 5 to 100 amino
26 acids;

27

28 expressing both constructs in the same cell
29 following co-transfection of a large population
30 of cells with both constructs;

31

1 measuring fluorescence produced in the cell,
2 selection of those cells with higher
3 fluorescence, using either a fluorescence
4 activated cell sorting machine or alternatively
5 by employing laser microdissection; and
6
7 clonally amplifying these fluorescent cells,
8 and sequencing the region of a large sample of
9 the constructs encoding the linkers and
10 determining the length of the linkers.

11
12 Preferably the linkers of the first and / or second
13 construct are comprised of flexible pentapeptide
14 sequences.

15
16 Preferably the pentapeptide is comprised of
17 substantially hydrophillic amino-acid residues.
18 More preferably the pentapeptide is a sequence such
19 as glycyl-glycyl-glycyl-glycyl-serine.

20
21 Preferably the number of peptapeptide sequences in
22 the linker is determined by sequencing.

23
24 Preferably the linker of the first and / or second
25 construct length is between 10 to 100 amino acids.
26 Alternatively the linker can be between 15 to 100
27 amino acids in length. In a yet further alternative
28 the linker can be 20 to 100 amino acids in length.
29 As a further alternative the linker can be 30 to 100
30 amino acids in length.

31

1 A distribution of occurrence of linker lengths will
2 be obtained in the fluorescent cells selected, with
3 a sharp cutoff at the lower limit reflecting the
4 minimum linker length capable of spanning the
5 separation of the fusion termini of the interacting
6 peptides and thus allowing productive association of
7 the fluorogenic fragments. A maximum value for this
8 distance may be evaluated in Ångstroms on the basis
9 that each amino acid residue contributes 3.7Å to the
10 length of each linker in an extended backbone
11 conformation.

12

13 Further assay methods of the present invention may
14 be used to detect the interactions of three or more
15 agents in a trimeric or higher order complex.

16

17 In one example, three fluorescent fragments may
18 be provided by introducing two split points as
19 discussed above into the fluorescent protein, each
20 fragment being fused to a peptide. On interaction
21 of the peptides the three or more fluorescent
22 fragments are brought together such that they can
23 functionally associate and generate a fluorescent
24 signal capable of being detected.

25

26 In another example one or more of the three
27 fluorescent fragments can be fused to a test agent
28 such as a small molecule, such as a metal ion. In
29 this manner, protein interactions which require the
30 participation of additional test agents, such as
31 small molecules can be detected.

32

1 Modulation of the interaction between peptides may
2 be a desirable outcome in the treatment of certain
3 clinical conditions, or as a research tool to study
4 peptide to peptide interactions. For example,
5 modulation of protein to protein interactions may
6 facilitate the task of determining the steps of
7 complex pathways by the provision of means to
8 promote or inhibit a specific interaction, allowing
9 the effects of other proteins to be studied in
10 better detail.

11

12 Many protein to protein interactions require the
13 participation of small molecules or peptides. Such
14 a requirement can be determined by simply adding
15 small molecule ligands or the peptides to the
16 components of the assay to determine if these
17 modulate protein to protein interaction as measured
18 by an alteration in fluorescent signal.

19

20 Thus in an eighth aspect there is provided an assay
21 for determining whether a candidate agent modulates
22 protein to protein interactions comprising the
23 steps:

24

25 providing a first construct encoding a first
26 fragment of fluorescent protein, a first bait
27 peptide and a linker portion of at least 5
28 amino acid residues interposed between the
29 first fragment and the bait peptide;

30

31 providing a second construct encoding a second
32 fragment of fluorescent protein which is

1 complementary to said first fluorescent
2 fragment, a prey peptide and a linker portion
3 of at least 5 amino-acids interposed between
4 the second fragment and the prey peptide;

5
6 providing a putative modulating agent;

7
8 expressing both constructs in the same cell;
9 and

10
11 measuring fluorescence produced in the cell in
12 the presence and absence of said putative
13 modulating agent

14
15 wherein a reduction in fluorescence in the
16 presence of said modulating agent compared to
17 fluorescence in the absence of said candidate
18 modulating agent is indicative of inhibition of
19 complex formation by the modulating agent and
20 an increase in fluorescence is indicative of
21 enhancement of complex formation by the
22 modulating agent.

23
24 Preferably the linker of the first and / or second
25 construct comprises a linker portion which encodes
26 between 15 to 100 amino acid residues.

27
28 Preferably the linker of the first and / or second
29 construct is comprised of substantially hydrophillic
30 amino-acid residues.

31

1 More preferably the linker of the first and / or
2 second construct is comprised of multiples of a
3 pentapeptide sequence such as glycyl-glycyl-glycyl-
4 glycyl-serine.

5
6 More preferably the linker of the first and / or
7 second construct is greater than 20 amino acids,
8 more preferably greater than 25 amino acids, more
9 preferably greater than 30 amino acids, more
10 preferably greater than 35 amino acids, even more
11 preferably greater than 40 amino acids, even more
12 preferably greater than 50 amino acids and yet more
13 preferably greater than 55 amino acids in length.

14
15 Preferably, the linker of the first and / or second
16 construct encodes up to 60 amino acids.

17
18 In a ninth aspect there is provided an assay for
19 determining whether a candidate agent modulates
20 protein to protein interactions comprising the
21 steps:

22
23 providing a first polypeptide comprising
24 a first fragment of fluorescent protein, a bait
25 peptide and a linker portion of at least 5
26 amino acid residues interposed between the
27 first fragment and the bait peptide;

28
29 providing a second polypeptide comprising a
30 second fragment of fluorescent protein which is
31 complementary to said first fluorescent
32 fragment, a prey peptide and a linker portion

1 of at least 5 amino-acids interposed between
2 the second fragment and the prey peptide;
3
4 providing a putative modulating agent; and
5
6 measuring fluorescence produced in the presence
7 and absence of said putative modulating agent
8
9 wherein a reduction in fluorescence in the
10 presence of said modulating agent compared to
11 fluorescence in the absence of said candidate
12 modulating agent is indicative of inhibition of
13 complex formation by the modulating agent and
14 an increase in fluorescence is indicative of
15 enhancement of complex formation by the
16 modulating agent.
17
18 Preferably the linker of the first and / or second
19 polypeptide comprises between 15 to 100 amino acid
20 residues.
21
22 Preferably the linker of the first and / or second
23 polypeptide is comprised of substantially
24 hydrophillic amino-acid residues.
25
26 More preferably the linker of the first and / or
27 second polypeptide is comprised of multiples of a
28 pentapeptide sequence such as glycyl-glycyl-glycyl-
29 glycyl-serine.
30
31 More preferably the linker of the first and / or
32 second polypeptide is greater than 20 amino acids,

1 more preferably greater than 25 amino acids, more
2 preferably greater than 30 amino acids, more
3 preferably greater than 35 amino acids, even more
4 preferably greater than 40 amino acids, even more
5 preferably greater than 50 amino acids and yet more
6 preferably greater than 55 amino acids in length.

7

8 Preferably, the linker of the first and / or second
9 construct polypeptide is up to 60 amino acids.

10

11 Thus the above assay can be used to select compounds
12 capable of triggering, stabilising or destabilising
13 peptide to peptide interactions.

14

15 As will be apparent, the assay of the present
16 invention can be applied in a format appropriate for
17 large scale screening, for example, combinatorial
18 technologies can be employed to construct
19 combinatorial libraries of small molecules or
20 peptides to test as modulating agents.

21

22 Preferably, structural information on the peptide to
23 peptide interaction to be modulated is obtained by
24 testing different agents to determine if they are
25 modulating agents.

26

27 For example, each of the interacting pair can be
28 expressed and purified and then allowed to interact
29 in suitable *in vitro* conditions. Optionally the
30 interacting peptides can be stabilised by
31 crosslinking or other techniques. The interacting
32 complex can be studied using various biophysical

1 techniques such as X-ray crystallography, NMR, or
2 mass spectrometry. In addition, information
3 concerning the interaction can be derived through
4 mutagenesis experiments for example alanine scanning,
5 or altering the charged amino acids or hydrophobic
6 residues on the exposed surface of the bait or prey
7 peptide being tested.

8
9 Based on the structural information obtained,
10 structural relationships between the interacting
11 peptides as well as between the modulating compound
12 and the interacting peptides can be elucidated.
13 Further, the three dimensional structure of the
14 interacting moieties and / or that of the modulating
15 compound can provide information to determine
16 suitable lead compounds able to modulate
17 interaction, which medicinal chemists can use to
18 design analog compounds having similar moieties and
19 structures.

20
21 In a tenth aspect, the invention provides compounds
22 obtainable by an assay of the invention, for example
23 small molecules, peptides or nucleic acids which
24 interact with the peptides being tested and modulate
25 the formation of a peptide complex.

26
27 Modulator compounds obtained accordingly to the
28 method of invention may be prepared as a
29 pharmaceutical preparation or composition.

30 Such preparations will comprise the modulating
31 compound and a suitable carrier, diluent or
32 excipient. These preparations may be administered

1 by a variety of routes, for example, oral, buccal,
2 topical, intramuscular, intravenous, subcutaneous or
3 the like.

4

5 According to an eleventh aspect of the present
6 invention there is provided a method of
7 manufacturing a composition or preparation
8 comprising:

9

10 performing an assay for determining whether a
11 candidate agent modulates peptide to peptide
12 interactions as described above; and

13

14 formulating said agent into a composition.

15

16 Also provided are nucleic acid constructs for use in
17 the invention.

18

19 Accordingly, in a twelfth aspect, there is provided
20 a nucleic acid construct encoding a fragment of a
21 fluorescent protein, a peptide and a linker portion
22 of at least 15 amino acid residues interposed
23 between said fragment and said peptide, wherein said
24 fragment of fluorescent protein is capable of
25 functional association with a complementary fragment
26 of fluorescent protein such that on functional
27 association of said fragments fluorescence is
28 enabled.

29

30 Preferably the first and / or second construct
31 comprises a linker portion which encodes between 15
32 to 100 amino acid residues.

1
2 Preferably the linker is comprised of substantially
3 hydrophillic amino-acid residues.
4
5 More preferably the linker is comprised of multiples
6 of a pentapeptide sequence such as glycyl-glycyl-
7 glycyl-glycyl-serine.
8
9 More preferably the linker portion encodes greater
10 than 20 amino acids, more preferably greater than 25
11 amino acids, more preferably greater than 30 amino
12 acids, more preferably greater than 35 amino acids,
13 even more preferably greater than 40 amino acids,
14 even more preferably greater than 50 amino acids and
15 yet more preferably greater than 55 amino acids in
16 length.
17
18 According to a thirteenth aspect of the invention
19 there is provided an expression vector comprising at
20 least one construct encoding a fragment of a
21 fluorescent protein, a peptide and a linker portion
22 of at least 15 amino acid residues interposed
23 between said fragment and said peptide, wherein said
24 fragment of fluorescent protein is capable of
25 functional association with a complementary fragment
26 of fluorescent protein such that on functional
27 association of said fragments fluorescence is
28 enabled operably linked to at least one regulatory
29 sequence for the expression of the construct.
30
31 The vector can be introduced into the cell using any
32 known techniques such as calcium phosphate

1 precipitation, lipofection, electroporation and the
2 like.

3
4 Where two vectors are provided, and each vector
5 encodes a different construct, for example a bait
6 construct and a prey construct, the vectors can be
7 transfected into the same cell or alternatively into
8 two different cells which are subsequently fused
9 together by cell fusion or other suitable
10 techniques.

11
12 In a fourteenth aspect of the invention there is
13 provided a cell transformed with a vector comprising
14 at least one construct encoding a fragment of a
15 fluorescent protein, a peptide and a linker portion
16 of at least 15 amino acid residues interposed
17 between said fragment and said peptide, wherein said
18 fragment of fluorescent protein is capable of
19 functional association with a complementary fragment
20 of fluorescent protein such that on functional
21 association of said fragments fluorescence is
22 enabled operably linked to at least one regulatory
23 sequence for the expression of the construct:

24
25 Cells which may be transformed include eukaryotic
26 cells, such as yeast, insect, plant, mammalian,
27 primate and human cells. Mammalian cells may be
28 primary cells or transformed cells, including tumour
29 cells. The system is not restricted to intracellular
30 (single cell) interactions. In multicellular
31 organisms amenable to genetic manipulation, a
32 protein-hapto-GFP construct could be released from

1 one cell or organ and be recognised by another
2 protein-(receptor)-haptoGFP fusion to indicate
3 localisation of filled receptors by the resultant
4 fluorescent signal.

5
6 In cell free systems such additional proteins as
7 required for expression may be included, for
8 example, by being provided by expression from
9 suitable recombinant expression vectors.

10
11 In addition, there is provided in a fifteenth aspect
12 of the invention a polypeptide encoded by a
13 construct encoding a fragment of a fluorescent
14 protein, a peptide and a linker portion of at least
15 15 amino acid residues interposed between said
16 fragment and said peptide, wherein said fragment of
17 fluorescent protein is capable of functional
18 association with a complementary fragment of
19 fluorescent protein such that on functional
20 association of said fragments fluorescence is
21 enabled.

22
23 In a sixteenth aspect of the invention there is
24 provided a library of polypeptides as encoded by
25 constructs according to the fifteenth aspect of the
26 invention.

27
28 Preferably the polypeptides of the library comprise
29 a linker portion which encodes between 15 to 100
30 amino acid residues.

31

1 Preferably the polypeptides of the library comprise
2 a linker of substantially hydrophillic amino-acid
3 residues.

4

5 More preferably the linker is comprised of multiples
6 of a pentapeptide sequence such as glycyl-glycyl-
7 glycyl-glycyl-serine.

8

9 More preferably the linker portion is greater than
10 20 amino acids, more preferably greater than 25
11 amino acids, more preferably greater than 30 amino
12 acids, more preferably greater than 35 amino acids,
13 even more preferably greater than 40 amino acids,
14 even more preferably greater than 50 amino acids and
15 yet more preferably greater than 55 amino acids in
16 length.

17

18 Preferably, the linker comprises up to 60 amino
19 acids.

20

21 Preferably the first and / or second linker of the
22 vector, or polypeptide can comprises between 15 to
23 100 amino acid residues.

24

25 Preferably the first and / or second linker is
26 comprised of substantially hydrophillic amino-acid
27 residues.

28

29 More preferably the first and / or second linker is
30 comprised of multiples of a pentapeptide sequence
31 such as glycyl-glycyl-glycyl-glycyl-serine.

32

1 More preferably the first and / or second linker
2 portion encodes or comprises greater than 20 amino
3 acids, more preferably greater than 25 amino acids,
4 more preferably greater than 30 amino acids, more
5 preferably greater than 35 amino acids, even more
6 preferably greater than 40 amino acids, even more
7 preferably greater than 50 amino acids and yet more
8 preferably greater than 55 amino acids in length.

9
10 According to a seventeenth aspect of the present
11 invention there is provided a kit comprising at
12 least two constructs according to the twelfth aspect
13 of the invention and means to express the
14 constructs.

15
16 The kit may further include test agents, which may
17 enhance or inhibit peptide to peptide interaction.

18
19 In another embodiment the kit includes cell lines in
20 which the vector of the thirteenth aspect can be
21 expressed.

22
23 Alternatively the kit can comprise at least one
24 polypeptide of the fifteenth aspect of the invention
25 and means for introducing the polypeptide into a
26 cell.

27
28 Additionally, ~~the kit can include~~ instructions for
29 using the kit to practise the present invention.

30 The instructions should be in writing in a tangible
31 form or stored in an electronically retrievable
32 form.

1

2 Brief description of the figures

3

4 The present invention will now be described with
5 reference to the following non-limiting examples and
6 with reference to the figures, wherein:

7

8 Figure 1a is a ribbon diagram of EGFP annotated
9 with split point sites;

10

11 Figure 1b is an illustration of the split
12 points and the related sequences surrounding
13 these split points of EGFP;

14

15 Figure 2 is a representation of a hapto-EGFP
16 with a 26 residue linker between the
17 fluorogenic fragments and the bait and prey
18 proteins respectively;

19

20 Figure 3 is a graph of the fluorescence
21 produced by the association of fragments joined
22 to linkers of different lengths, (A) Cells
23 cotransfected with pN157(6)zip and pzip(4)C158
24 in which a functional leucine zippers mediate
25 the association of haptoEGFP1-157 and
26 haptoEGFP158-238 to generate fluorescence, (B)
27 Negative control cotransfection using pN157(6)
28 and p(4)C158 which lack sequences encoding the
29 leucine zippers and as such fail to generate
30 fluorescence, (D) Cells cotransfected with
31 pN172(6)zip and pzip(4)C173 in which a
32 functional leucine zipper mediated association

1 of haptoEGFP1-172 and haptoEGFP173-238 occurs
2 to generate fluorescence which is of greater
3 intensity to that observed with the 157/158
4 split point (E) Negative control
5 cotransfection using pN172(6) and p(4)C173
6 which lack sequences encoding the leucine
7 zippers and as such fail to generate
8 fluorescence, (C and F) Confocal images of
9 cotransfected cells from (A) and (D) showing
10 the intracellular localisation of fluorescence.
11 Vero cells were cotransfected with plasmids
12 encoding linkers ranging in length from 4 to 26
13 amino acids and UV images were collected at 24
14 hours post-transfection using identical
15 exposure times, (G) pN157(6)zip and
16 pzip(4)C158 (H) pN157(16)zip and pzip(14)C158
17 (I) pN157(26)zip and pzip(24)C158 (J)
18 pN157(26)zip and pzip(4)C158 (K) pN157(6)zip
19 and pzip(24)C158 (L) a negative untransfected
20 control illustrates the background fluorescence
21 level, *Italicised figures in brackets indicate*
22 *the length of the hydrophilic linker;*
23

24 Figure 4 shows the importance of being able to
25 fuse the interacting peptide to either the N,
26 N', C or C' of the fluorescent fragment.

27

28 Structural studies of GFP have revealed that the
29 protein exists as a compact cylindrical structure,
30 with eleven beta-sheet strands forming the walls of
31 the cylinder, the N and C termini being at close
32 proximity at the base of the structure. Sections of

1 alpha-helix form caps on the end of the cylinders
2 and an irregular alpha-helical segment also provides
3 a scaffold for the fluorophore which is located in
4 the geometric center of the cylinder. This folding
5 motif, with beta-sheet outside and helix inside is
6 known as beta-can.

7
8 The inventors have shown that fluorescence can be
9 generated following functional association of two
10 separate fragments of GFP molecules (haptoGFPs) when
11 driven by the interactions of a pair of proteins
12 fused both to the new C' and N' termini of each
13 haptoGFP and also to the existing termini.

14
15 Functional association of fragments of fluorescent
16 proteins, brought together by the interaction of
17 peptides fused to the fragments, to screen for
18 protein to protein interactions requires that the
19 fragments reliably functionally associate only after
20 interaction of the fused peptides.

21
22 Reliable functional association has to date not been
23 obtainable due to the possibility of steric
24 hindrance and steric constraints on the functional
25 association of haptoGFPs when bulky proteins are
26 associated to the GFP fragments.

27
28 To overcome problems of steric hindrance, linker
29 regions of at least 15 residues are provided between
30 the peptide being tested for interaction and the
31 associated fluorogenic fragment. This provides the
32 peptide with considerable flexibility relative to

1 the fluorogenic fragment to bind to another peptide
2 being tested while still enabling the fluorogenic
3 fragments to complement each other and cause
4 detectable fluorescence to be generated.

5
6 To prepare GFP fragments, which are capable of
7 functional association, split points were generated
8 at various points along the 239 residue length of
9 the GFP protein, resulting in the generation of new
10 C' and N' termini which, in three dimensions, are
11 located at the top and at the base of the beta-can
12 structure.

13
14 Split points were introduced based on a structure
15 driven approach between hydrophilic residues. The
16 eleven strands of the beta structure making up the
17 beta-can topology of EGFP are characterised by
18 forming three instances of a tripartite antiparallel
19 sheet motif joined edge to edge around the periphery
20 of the 'can'; with the remaining two beta strands
21 completing the cylindrical structure. The most
22 successful split points obtained to date occur in
23 the third tripartite motif between hydrophilic
24 residues allowing adjacent hydrophobic side chains
25 to promote refolding of the haptogFPs.

26
27 As shown in the non exhaustive list of Table 1 a
28 number of split points were identified using the
29 above approach. It would appear that each split
30 point in Table 1 is simply one example of a range of
31 potentially useful split points, the range being
32 shown in parentheses of Table 1.

Table 1

Split point Number	Residue position in EGFP	Possible range
1	23/24	(23-25)
2	38/39	(36-41)
3	50/51	(48-54)
4	76/77	(75-91)
5	89/90	(75-90)
6	102/103	(101-103)
7	116/117	(115-118)
8	132/133	(129-143)
9	142/143	(129-143)
10	157/158	(155-160)
11	172/173	(171-175)
12	190/191	(187-199)
13	211/212	(207-218)
14	214/215	(207-218)

To extend the versatility of the hapto-EGFP method, constructs were created where instead of using C' and N' for the attachment of heterologous proteins, the endogenous termini, N or C together with one of the new N' or C' termini were used. Using this technique the bait and prey peptides can be added such that they are orientated to the associated fluorogenic fragments in the same direction as each other, for example both attached to bottom of the β -can structure of GFP or in the opposite direction, for example the bait peptide is attached to the

1 bottom of the β -can structure of GFP, while the prey
2 protein is attached to the top of the β -can
3 structure of GFP. As will be understood by those
4 skilled in the art, and as shown in figures 4 A & B,
5 as peptides interact with each other in a particular
6 orientation, then the direction of the linkage of
7 the peptide to the N, N', C or C' end of the
8 fluorogenic fragment becomes important in certain
9 circumstances so as to allow the fluorescent protein
10 fragments to functionally interact following
11 interaction of the peptides.

12
13 Thus, to minimise interference with the refolding
14 and association of the two hapto-EGFPs during
15 assembly, it would appear that the most versatile
16 split points may occur at the bottom of the β -can.
17 These effects may be minimised by the use of longer
18 linkers to accommodate adverse topology (Figure 4C).

19

20 Example 1

21

22 As shown in figures 2 and 3, hapto-EGFP with a 26-
23 residue linker between the fluorogenic fragments and
24 the bait and prey proteins respectively were
25 produced without loss of fluorescence. These
26 linkers may be lengthened using overlapping
27 oligonucleotides encoding repeating (GGGS)_x units.
28 This was achieved by using unique Sac I and BamHI
29 restriction sites present in the core expression
30 vectors pN^{EGFP}(Sac)zip and pzip(Bam)C^{EGFP}.

31

1 To test whether it was possible to obtain
2 fluorescence when endogenous N or C termini and a
3 new N' or C' terminus are used to attach heterologous
4 proteins the fusion (F) and haemagglutinin (H)
5 membrane proteins of measles virus (MV) were used.
6

7 Measles virus (MV) infection is mediated by a
8 complex of two viral envelope proteins,
9 haemagglutinin (H) glycoprotein and fusion (F)
10 glycoprotein that bind to each other and then
11 complex with surface receptors to aid the fusion of
12 the virus with the plasma membrane. The H
13 glycoprotein is dimerised in the endoplasmic
14 reticulum and is thought to exist on the cell
15 surface as a tetramer (dimer of dimers). The fusion
16 (F) glycoprotein, is synthesised as an inactive
17 precursor (F_0) which is a highly conserved type I
18 transmembrane glycoprotein of about 60kDa, which is
19 cleaved by furin in the trans-golgi to yield the
20 41kDa (f_1) and the 18kDa (f_2) disulphide-linked
21 activated F-protein. Infection of the measles virus
22 is dependent on the interaction of the F/H complex
23 with cell surface receptors.
24

25 A pair of constructs was generated which encoded the
26 H glycoprotein fused at its N terminus to either the
27 C' terminal residue of hapto-EGFP¹⁻¹⁵⁷ ($N^{1/157}$) in the
28 first member, or to the natural C terminal residue
29 of the complimentary hapto-EGFP¹⁵⁸⁻²³⁹ (C158/239) in
30 the second member of the pair. Each construct
31 includes an encoded linker between these two
32 proteins.

1 Linkers generated using overlapping oligonucleotides
2 which contain *Sfi* IA and *Sfi* IB restriction sites
3 were introduced into pN^{1/157}(16)zip and pC^{158/239}(14)zip
4 constructs. The complete H gene ORF was amplified
5 by polymerase chain reaction (PCR) using primers
6 containing *Sfi* IA and *Sfi* IB restriction sites and
7 the PCR products used to generate pN^{1/157}(16)MV-H and
8 pC^{158/239}(14)MV-H. A similar pair of constructs
9 employing N and N' hapto-EGFP fusions, pMV-F(16)N^{1/157}
10 and pMV-F(14)C^{158/238} could be generated from existing
11 clones using a similar strategy.

12
13 Vero cells (African green monkey kidney-derived cell
14 line) were transiently transfected with pN^{1/157}(16)MV-
15 H and pMV-F(14)C^{158/238} constructs, the proteins
16 expressed and phase contrast microscopy used to
17 determine whether the modified glycoproteins
18 retained their fusogenicity.

19
20 Real-time observation by ultraviolet and confocal
21 microscopy indicated if fluorescence was generated
22 upon expression of the hapto-EGFP/glycoprotein
23 fusions.

24
25 As cells transiently transfected with both F and H
26 expression plasmids form syncytia in the absence of
27 viral replication, the formation of syncytia can be
28 used to assay for successful transfection of both
29 plasmids.

30
31 The size of the syncytia was compared with controls
32 to establish whether transfection had occurred. UV

1 and confocal microscopy were used to examine the
2 fluorescence so as to verify that association
3 between H protein oligomers and F proteins had taken
4 place. Confocal microscopy and image reconstruction
5 were also used to determine the intracellular
6 localisation of H protein oligomers during formation
7 of the fusion complex.

8
9 Using the above vectors the intracellular
10 association of F and H proteins and their
11 trafficking from the endoplasmic reticulum (ER) to
12 the plasma membrane was tracked. Further, membrane
13 receptor proteins which interact with the H protein
14 could be identified as could cytoplasmic proteins
15 which interact with known MV receptors and which may
16 therefore initiate downstream signalling events.

17
18 Example 2

19
20 The above constructs could also be incorporated into
21 a recombinant measles viral genome and the
22 experiments repeated to determine if the above
23 constructs could be used in *in vivo* viral studies,

24
25 The type-I F glycoprotein is proposed to form
26 trimers.

27
28 In this example two split points could be introduced
29 into the EGFP. The constructs pMV-F(16)N^{1/157}, pMV-
30 F(14)M^{158/190} and pMV-F(14)C^{191/239} could be generated.

31

1 The method could then be adapted to screen for and
2 identify virus receptors.

3

4 This could be tested with MV and applied to the
5 closely related mumps virus (MuV).

6

7 Example 3

8

9 Fusion of oligonucleotides encoding hapto-EGFP
10 sequences to members of a cDNA library.

11

12 Firstly, the sequence encoding the hapto-EGFP may be
13 fused to the 5' end of the library due to the
14 presence of downstream stop codons in the cDNA.

15

16 Secondly, constructs are required to be generated
17 for all three reading frames to ensure that one is
18 in the correct reading frame.

19

20 Thirdly, the cDNA sequences are required to be
21 obtained from a source which permits directional
22 cloning into restriction sites which are extremely
23 rare in mammalian DNA. Such sequences are to be
24 found in the *Large-Insert cDNA library* (Clontech).

25

26 A core panning vector could be engineered from
27 existing plasmids to contain a CMV promoter, an
28 initiation codon and sequences encoding a hapto-EGFP
29 and an intervening linker, an *Sfi* IA site and an *Sfi*
30 IB site, a stop codon and an SV40 polyadenylation
31 signal. Two additional screening vectors could be
32 generated to include one and two extra nucleotides

1 between the linker and the *Sfi* IA site to correct
2 the reading frame. cDNA fragments, flanked with *Sfi*
3 IA and *Sfi* IB sites obtained from the library could
4 be cloned downstream of the optimised hapto-EGFP
5 linker constructs. The hapto-EGFP library could then
6 be transfected into CHO cells and a mixed population
7 of cells selected using G418 and passaged to
8 confluency. These cells could then be transfected
9 with CD46-haptoEGFP or the equivalent SLAM plasmid.

10

11 Where interaction between the peptides being
12 screened occurs, fluorescence is generated.

13

14 Any cells which fluoresce can then be isolated by
15 fluorescent laser microdissection and single cell
16 RT-PCR performed to identify mRNA which encodes
17 peptides which interact with the cytoplasmic tails
18 of the receptor molecules.

19

20 Although the invention has been particularly shown
21 and described with reference to particular examples,
22 it will be understood by those skilled in the art
23 that various changes in the form and details may be
24 made therein without departing from the scope of the
25 present invention.

26

27

28

1 Claims

2

3 1. A protein interaction system said system
4 comprising a first construct which encodes a
5 first fragment of fluorescent protein, a first
6 bait peptide and a linker portion encoding at
7 least 5 amino acid residues interposed between
8 the first fragment and the bait peptide and a
9 plurality of second constructs encoding a
10 second fragment of fluorescent protein, a prey
11 peptide and a linker portion encoding at least
12 5 amino acid residues interposed between the
13 second fragment and the prey peptide and on
14 interaction of the bait and a prey peptide the
15 first and second fragments of the fluorescent
16 protein complement each other such that
17 functional association of the first and second
18 fragments promotes fluorescence, wherein at
19 least two of the prey proteins have different
20 amino acid sequences.

21

22 2. A protein interaction system as claimed in
23 claim 1 wherein all the prey proteins have
24 different amino acid sequences.

25

26 3. A protein interaction system as claimed in
27 claim 1 or claim 2 wherein the first and / or
28 second construct comprises a linker portion
29 which encodes between 15 to 100 amino acid
30 residues.

31

32

- 1
2 4. A library of constructs encoding a fragment of
3 fluorescent protein, a peptide and a linker
4 portion of at least 5 amino acids interposed
5 between said fragment and peptide wherein said
6 fragment of fluorescent protein is capable of
7 functional association with a complementary
8 fragment of fluorescent protein such that on
9 functional association of said fragments
10 fluorescence is enabled wherein the library of
11 constructs encodes a plurality of different
12 peptides.
13
- 14 5. A library of polypeptides, each polypeptide
15 comprising a fragment of fluorescent protein, a
16 peptide and a linker portion of at least 5
17 amino acid residues interposed between the
18 fragment and the peptide of the polypeptide
19 wherein said fragment of fluorescent protein is
20 capable of functional association with a
21 complementary fragment of fluorescent protein
22 such that on functional association of said
23 fragments fluorescence is enabled wherein the
24 library comprises a plurality of different
25 peptides.
26
- 27 6. A protein interaction monitoring system, said
28 system comprising a first polypeptide
29 comprising a first fragment of fluorescent
30 protein, a bait peptide and a linker portion of
31 at least 5 amino acid residues interposed
32 between the first fragment and the bait peptide

1 and a plurality of second polypeptides
2 comprising a second fragment of fluorescent
3 protein, a prey peptide and a linker portion of
4 at least 5 amino acid residues interposed
5 between the second fragment and the prey
6 peptide and on interaction of the bait and a
7 prey peptide the first and second fragments of
8 the fluorescent protein complement each other
9 such that functional association of the first
10 and second fragments promotes fluorescence,
11 wherein at least two of the prey proteins have
12 different amino acid sequences.

13

14 7. An assay method to determine peptide to peptide
15 interactions comprising the steps of:

16

17 providing a first construct, said construct
18 encoding a first fragment of fluorescent
19 protein, a first bait peptide and a linker
20 portion of at least 5 amino acid residues
21 interposed between the first fragment and the
22 bait peptide;

23

24 providing a plurality of second constructs said
25 constructs encoding a second complementary
26 fragment of fluorescent protein, a prey peptide
27 and a linker portion of at least 5 amino acids
28 interposed between the second fragment and the
29 prey peptide wherein at least two constructs
30 encode different prey proteins;

31

1 expressing both constructs in the same cell;
2 and
3
4 detecting fluorescence produced in the cell.
5
6 8. An assay method as claimed in claim 7 wherein
7 all the second constructs encode different prey
8 proteins.
9
10 9. An assay to determine peptide to peptide
11 interactions comprising the steps of:
12
13 providing a first polypeptide comprising a
14 first fragment of fluorescent protein, a first
15 bait peptide and a linker portion of at least 5
16 amino acid residues interposed between the
17 first fragment and the bait peptide;
18
19 providing a plurality of second polypeptides
20 comprising a second fragment of fluorescent
21 protein which is complementary to the first
22 fragment of fluorescent protein, a prey peptide
23 and a linker portion of at least 5 amino acids
24 interposed between the second fragment and the
25 prey peptide wherein at least two second
26 polypeptides encode different prey proteins;
27
28 mixing the first polypeptide and second
29 polypeptide together; and
30
31 detecting whether fluorescence is produced.
32

- 1 10. An assay which comprises the steps of providing
2 a first construct encoding a polypeptide
3 comprising a first fragment of fluorescent
4 protein, a first bait peptide and a linker
5 portion of at least 5 amino acid residues
6 interposed between the first fluorogenic
7 fragment and the first bait peptide:
8
9 providing a second construct encoding a
10 polypeptide comprising a second fragment of
11 fluorescent protein which is complementary to
12 said first fluorescent fragment, a second prey
13 peptide and a linker portion interposed between
14 the second fluorogenic fragment and the second
15 prey peptide;
16
17 causing the expression of both constructs
18 within the same living cell; and
19
20 observing the level of fluorescence produced
21 and its subcellular location in the cell at a
22 range of time points following co-expression of
23 both constructs.
24
- 25 11. An assay for estimating the maximum possible
26 separation of the fusion termini of the
27 interacting peptides:
28
29 providing a first construct encoding a first
30 fragment of fluorescent protein, a first bait
31 peptide and a linker portion of at least 5

1 amino acid residues interposed between the
2 first fragment and the bait peptide;
3
4 providing a second construct encoding a second
5 fragment of fluorescent protein which is
6 complementary to said first fluorescent
7 fragment, a prey peptide and a library of
8 linkers of lengths ranging from 5 to 100 amino
9 acids;
10
11 expressing both constructs in the same cell
12 following co-transfection of a large population
13 of cells with both constructs;
14
15 measuring fluorescence produced in the cell,
16 selection of those cells with higher
17 fluorescence, using either a fluorescence
18 activated cell sorting machine or alternatively
19 by employing laser microdissection; and
20
21 clonally amplifying these fluorescent cells,
22 and sequencing the region of a large sample of
23 the constructs encoding the linkers and
24 determining the length of the linkers.
25
26 12. An assay for determining whether a candidate
27 agent modulates peptide to peptide interactions
28 comprising the steps:
29
30 providing a first construct encoding a first
31 fragment of fluorescent protein, a first bait
32 peptide and a linker portion of at least 5

1 amino acid residues interposed between the
2 first fragment and the bait peptide;
3
4 providing a second construct encoding a second
5 fragment of fluorescent protein which is
6 complementary to said first fluorescent
7 fragment, a prey peptide and a linker portion
8 of at least 5 amino-acids interposed between
9 the second fragment and the prey peptide;
10
11 providing a putative modulating agent;
12
13 expressing both constructs in the same cell;
14 and
15
16 measuring fluorescence produced in the cell in
17 the presence and absence of said putative
18 modulating agent
19
20 wherein a reduction in fluorescence in the
21 presence of said modulating agent compared to
22 fluorescence in the absence of said candidate
23 modulating agent is indicative of inhibition of
24 complex formation by the modulating agent and
25 an increase in fluorescence is indicative of
26 enhancement of complex formation by the
27 modulating agent.
28
29 13. An assay for determining whether a candidate
30 agent modulates peptide to peptide interactions
31 comprising the steps:
32

1 providing a first polypeptide comprising
2 a first fragment of fluorescent protein, a bait
3 peptide and a linker portion of at least 5
4 amino acid residues interposed between the
5 first fragment and the bait peptide;

6
7 providing a second polypeptide comprising a
8 second fragment of fluorescent protein which is
9 complementary to said first fluorescent
10 fragment, a prey peptide and a linker portion
11 of at least 5 amino-acids interposed between
12 the second fragment and the prey peptide;

13
14 providing a putative modulating agent; and

15
16 measuring fluorescence produced in the presence
17 and absence of said putative modulating agent

18
19 wherein a reduction in fluorescence in the
20 presence of said modulating agent compared to
21 fluorescence in the absence of said candidate
22 modulating agent is indicative of inhibition of
23 complex formation by the modulating agent and
24 an increase in fluorescence is indicative of
25 enhancement of complex formation by the
26 modulating agent.

27

28 14. A method of manufacturing a composition or
29 preparation comprising:

30

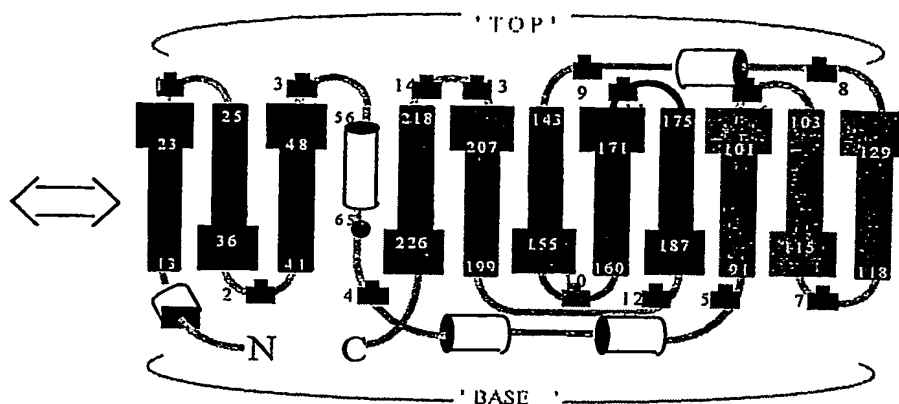
31 performing an assay as claimed in claim 11 or
32 claim 12 for determining whether a candidate

- 1 agent modulates peptide to peptide interactions
2 as described above; and
3
4 formulating said agent into a composition.
5
- 6 15. A nucleic acid construct encoding a fragment of
7 a fluorescent protein, a peptide and a linker
8 portion of at least 15 amino acid residues
9 interposed between said fragment and said
10 peptide, wherein said fragment of fluorescent
11 protein is capable of functional association
12 with a complementary fragment of fluorescent
13 protein such that on functional association of
14 said fragments fluorescence is enabled.
15
- 16 16. An expression vector comprising at least one
17 construct encoding a fragment of a fluorescent
18 protein, a peptide and a linker portion of at
19 least 15 amino acid residues interposed between
20 said fragment and said peptide, wherein said
21 fragment of fluorescent protein is capable of
22 functional association with a complementary
23 fragment of fluorescent protein such that on
24 functional association of said fragments
25 fluorescence is enabled wherein the construct
26 is operably linked to at least one regulatory
27 sequence for the expression of the construct.
28
- 29 17. A cell transformed with a vector as claimed by
30 claim 16.
31

- 1 18. A polypeptide encoded by a construct as claimed
2 in claim 15.
3
- 4 19. A library of polypeptides as encoded by
5 constructs of claim 18 wherein the constructs
6 encode at least two different prey proteins.
7
- 8 20. A kit comprising at least one pair of
9 constructs according to claim 15 wherein the
10 fragment of fluorescent protein of one member
11 of the pair is capable of functional
12 association with a complementary fragment of
13 the other member of the pair and means to
14 express the constructs.
15
- 16 21. A kit as claimed in claim 20 which further
17 comprises test agents, which enhance or inhibit
18 peptide to peptide interaction.
19
- 20 22. A kit as claimed in any one of claims 20 or 21
21 which further includes a vector of claim 16 and
22 cell lines in which the same can be expressed.
23
- 24 23. A kit as claimed in any one of claims 2 to 22
25 comprising at least one polypeptide of claim
26 18 and means for introducing the polypeptide
27 into a cell.
28
29

Figure I

A



GFP fold: The ribbon diagram to the left is coloured similarly to the cartoon on the right. β -sheets are indicated by arrows, α -helices by cylinders. Numbers within these symbols refer to sequence positions in EGFP (numbered according to the crystallographic structure - accession: 1emb). In the cartoon, connecting loops are shown by lines. Potential split points are starred and numbered sequentially from the N-terminus, (see **B** below for precise definitions). The fluorophore is represented by a green circle.

B

Possible split points in EGFP considered for hapto formation.

No.	Position	Top/ Bottom	> -----Sequence----- >
1	23/24	T	...D V N G ⁺ H K F S...
2	38/39	B	...G D A T Y ⁺ G K L...
3	50/51	T	...I C T T G ⁺ K L P...
4	76/77	B	...R Y P D H ⁺ M K Q...
5	89/90	B	...S A M P E ⁺ G Y V...
6	102/103	T	...F F K D D ⁺ G N Y...
7	116/117	B	...K F E G D ⁺ T L Y...
8	132/133	T	...D F K E D ⁺ G N I...
9	142/143	T	...H K L E Y ⁺ N Y N...
10	157/158	B	...A D K Q K ⁺ N G I...
11	172/173	T	...H N I E D ⁺ G S V...
12	190/191	B	...P I G D G ⁺ P V L...
13	211/212	T	...S K D P N ⁺ E K R...
14	214/215	T	...P N E K R ⁺ D H M...

Key: Residues at the new, internal C- and N-termini (C' & N') are shown in bold with ⁺ between them.

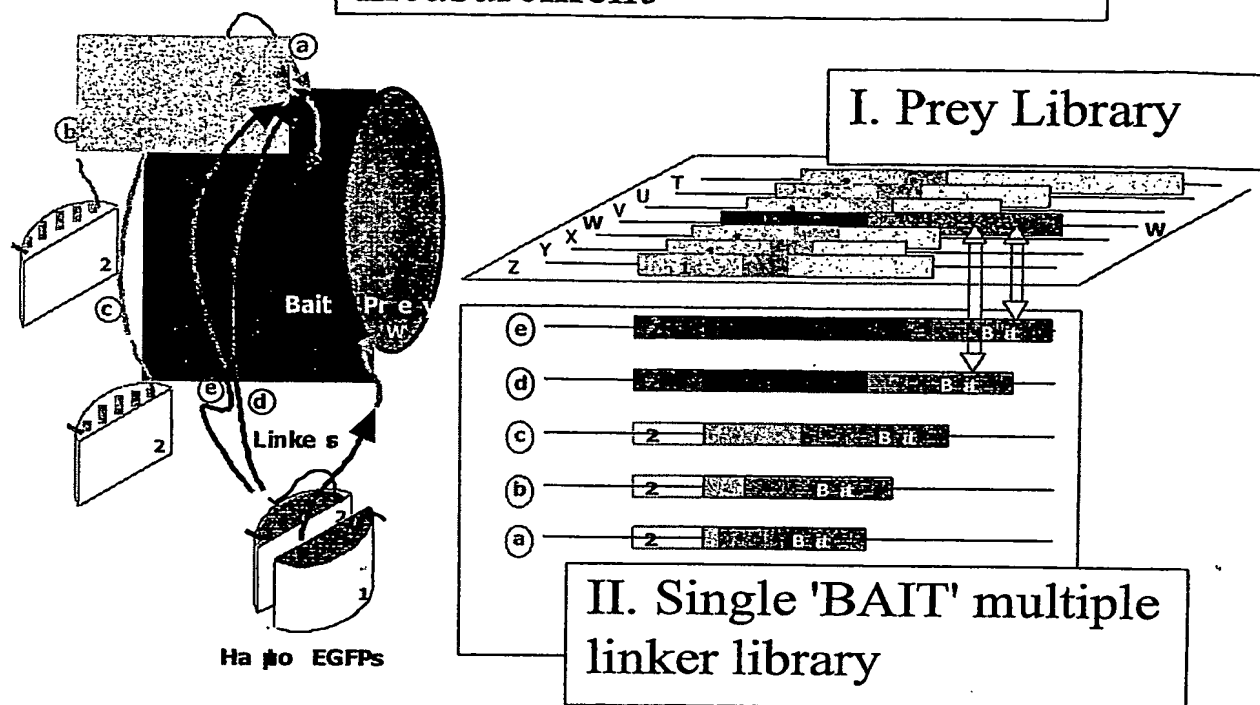
Adjacent hydrophobic residues are in *italics*.

The colours in the vertical bar correspond to the structural motifs of the cartoon.

2/4

Figure 2

Library search and proximity measurement

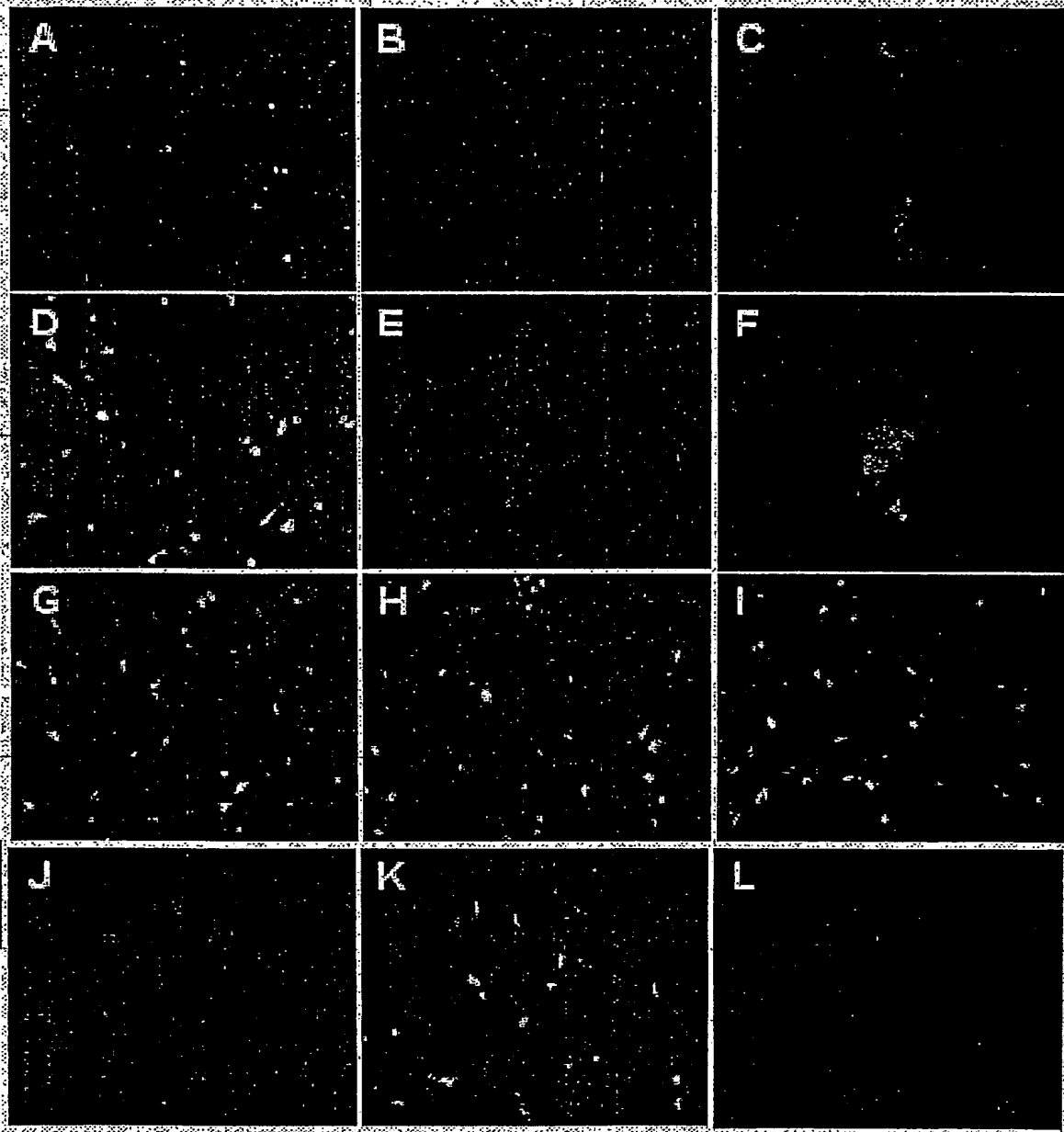


Schematic for protein to protein interaction searches by library interrogation. The two proteins in question are designated 'Bait' and 'W'. Two libraries are generated (I and II), one series of constructs (here designated T...Z, library I, >10,000 members) encodes a hapto-EGFP followed by a DNA sequence encoding a 60 residue linker attached to the 5'-end of a cDNA library, which contains the gene encoding the 'prey', "W" here. The second series of constructs (a...e here, library II, <20 members) encodes the complementary hapto-EGFP followed by a degenerate linker DNA sequence and the 'bait' gene. The individual components of the system are colour coded: blue - 'Bait'/'Prey'; pink - Linker; green - hapto-EGFP. All arrows indicate the direction of the polypeptide backbone (N→C)

A. 'Prey' identification: co-transfection with the 'prey' library (I) and construct 'e' (long linker - preferably 60 amino acid residues) from the 'bait' library (II) will generate fluorescent cells when the recipient cell receives a vector from library (I) bearing the 'W' gene (in this case) and a second vector bearing the 'e' bait construct. Clonal expansion of these fluorescent cells allows identification of gene 'W'.

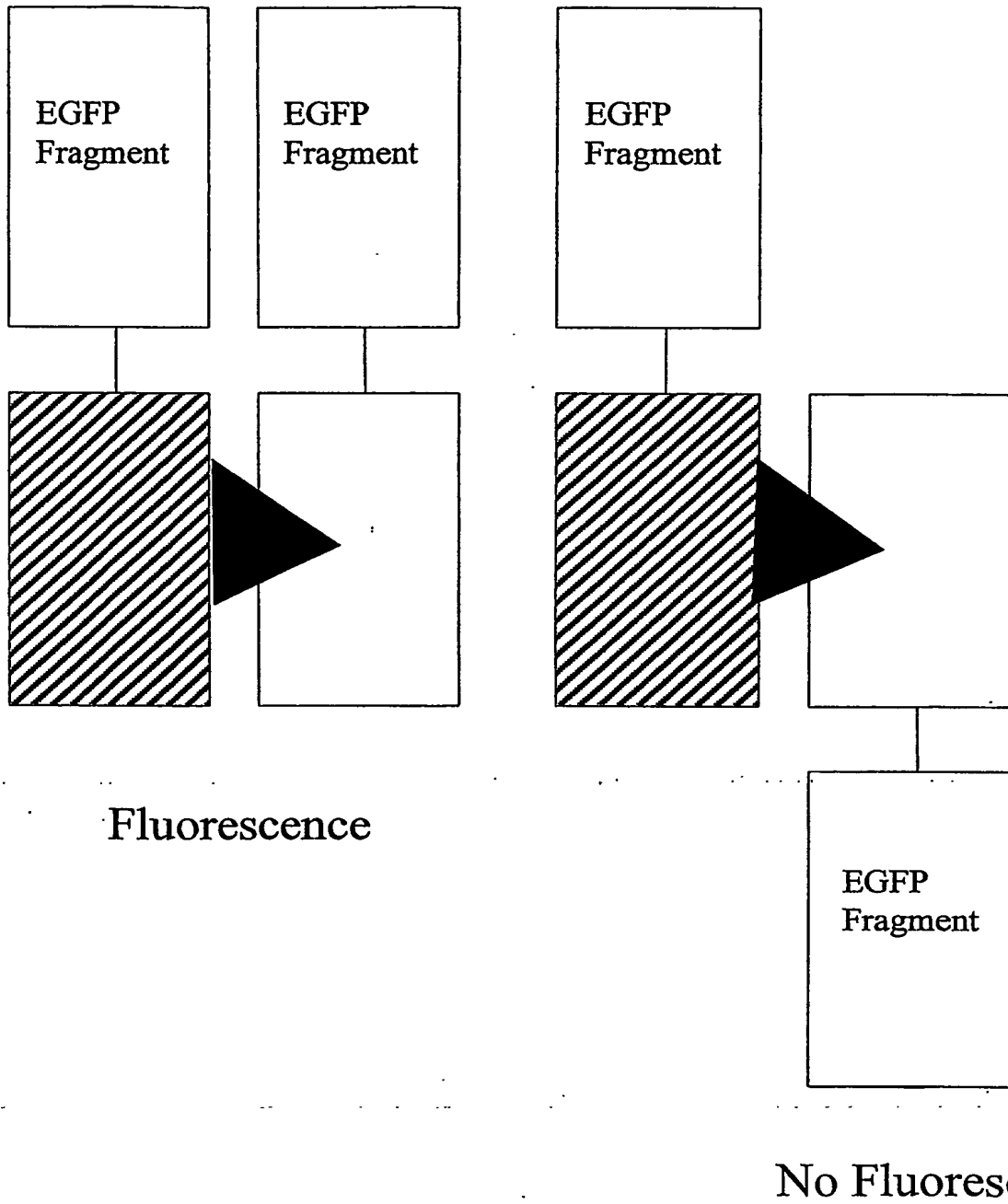
B. Proximity measurement: The clone(s) from step A are co-transfected with the 'bait' library (II). In this case cells showing fluorescence synthesise interacting proteins with a sufficiently long linker to allow productive complementary hapto-GFP interaction. ('d' or 'e' in this case), as shown to the left of the diagram. The hollow blue arrows in the right hand part of the diagram are intended to indicate that the interaction of the gene products with these two constructs will generate fluorescence, while other interactions between the product of gene 'W' and the bait protein will not give rise to fluorescent cells due to insufficient length of linker.

Figure 3



Fluorescent images of Vero cells transiently cotransfected with haptoEGFP expression constructs:

Figure 4



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record.**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.